Faculty of Cognitive Sciences and Human Development

REAL-TIME OPEN SET FACIAL RECOGNITION SYSTEM

Lim Yoong Kang

Bachelor of Science with Honours
(Cognitive Science)
2017

UNIVERSITI MALAYSIA SARAWAK

Grade: A

Please tick one

Final Year Project Report ☒

Masters ☐

PhD ☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the 13 day of JUNE year 2017.

Student's Declaration:

I, LIM YOONG KANG, 47408, FACULTY OF COGNITIVE SCIENCES AND HUMAN DEVELOPMENT, hereby declare that the work entitled, REAL-TIME OPEN SET RECOGNITION SYSTEM is my original work. I have not copied from any other students' work or from any other sources with the exception where due reference or acknowledgement is made explicitly in the text, nor has any part of the work been written for me by another person.

13 JUNE 2017

Yoong Kang

LIM YOONG KANG (47408)

Supervisor's Declaration:

I, ASSOCIATE PROFESSOR DR. TEH CHEE SIONG , hereby certify that the work entitled, REAL-TIME OPEN SET RECOGNITION SYSTEM was prepared by the aforementioned or above mentioned student, and was submitted to the "FACULTY OF COGNITIVE SCIENCES AND HUMAN DEVELOPMENT" as a *partial/full fulfillment for the conferment of BACHELOR OF SCIENCE WITH HONOURS (COGNITIVE SCIENCE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: _____ Date: 13 JUNE 2017

(ASSOC. PROF. DR. TEH CHEE SIONG)

I declare this Project/Thesis is classified as (Please tick (√)):

☐	**CONFIDENTIAL**	(Contains confidential information under the Official Secret Act 1972)*
☐	**RESTRICTED**	(Contains restricted information as specified by the organisation where research was done) *
☒	**OPEN ACCESS**	

I declare this Project/Thesis is to be submitted to the Centre for Academic Information Services (CAIS) and uploaded into UNIMAS Institutional Repository (UNIMAS IR) (Please tick (√)):

☒ **YES**
☐ **NO**

Validation of Project/Thesis

I hereby duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic and research purposes only and not for other purposes.
- The Centre for Academic Information Services has the lawful right to digitize the content to be uploaded into Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis if required for use by other parties for academic purposes or by other Higher Learning Institutes.
- No dispute or any claim shall arise from the student himself / herself neither a third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student himself/herself without first obtaining approval from UNIMAS.

Student's signature: _____ Supervisor's signature: _____

Date: 13 JUNE 2017 Date: 13 JUNE 2017

Current Address:
Jalan Datuk Mohammad Musa,
94300 Kota Samarahan, Sarawak, Malaysia

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the date of restriction indicated, and the reasons for the confidentiality and restriction.

REAL-TIME OPEN SET FACIAL RECOGNITION SYSTEM

LIM YOONG KANG

This project is submitted
in partial fulfillment of the requirements for a
Bachelor of Science with Honours
(Cognitive Science)

Faculty of Cognitive Sciences and Human Development
UNIVERSITI MALAYSIA SARAWAK
(2017)

The project entitled 'Real-Time Open Set Recognition System' was prepared by Lim Yoong Kang and submitted to the Faculty of Cognitive Sciences and Human Development in partial fulfillment of the requirements for a Bachelor of Science with Honours (Cognitive Science)

Received for examination by:

(ASSOC. PROF. DR. TEH CHEE SIONG)

Date:
13 June 2017

Grade
A

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God for giving me the power to believe in my passion and pursue my dreams. Thanks for granting me the wisdom and strengths on this academic journey through the end.

My deepest appreciation to my supervisor, Assoc. Prof. Dr. Teh Chee Siong for his patience, guidance, practical comments and constructive criticism throughout the process of the thesis. His guidance helps had made this thesis better than it started out. I am indebted to him for sharing expertise, and immense knowledge and encouragement extended to me.

Besides, I would like to special thanks to my parents, Mr. Lim Sng Fook and Mrs. Tai Lee Peng, who always gave me a lot of love, encouragement, and support throughout writing this thesis and my life in general. Besides, I would like to thank my friends who always help me.

Also, I would like to extend my gratitude to all the lecturers and staff that involved either directly or indirectly for the success of my research. Lastly, I would like to special thanks to everyone who have offered their help, comments, and advice for the improvement of this thesis

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Despite the world nature is dynamic and open, most of the recognition systems presume that the world is static and using a closed world model which treat every object must be known class in prior. However, in face recognition, we do not have the entire set of all possible faces. This project formularizes a generalized method for real-time recognition of open set recognition approach. The proposed approach enables the model to recognize an infinite set of faces in a myriad of unknown faces and unknown, unseen or new faces. The capability of constantly recognize unknown in real-time are the highlight of this system as an unknown face are treated as a valid outcome. This system utilized Viola-Paul algorithm to detect faces, PCA and LDA to extract the most significant data to describe the face, and Approximate Nearest Neighbour technique to recognized the previous known face and identify unknown faces via distance metric. Once unknown faces are identified, it will be learned and be labeled. Hence when the faces appear the second time, it will be recognized. In result, 60% accuracy is archived after fine tune the hyperparameter of feature selector and the distance metric threshold.

Keywords: real-time face detection, face feature extraction, approximate nearest neighbor, open set face recognition

ABSTRAK

Walaupun sifat dunia adalah dinamik dan terbuka, kebanyakan sistem pengecaman muka manusia menganggap bahawa dunia adalah statik dan menggunakan model dunia tertutup dengan menganggap setiap objek perlu diketahui kelasnya terlebih dahulu. Walau bagaimanapun, dalam pengecaman muka, kita tidak mempunyai seluruh set muka. Projek ini menggunakan kaedah umum untuk pengenalan muka dalam masa sebenar secara set terbuka. Pendekatan yang dicadangkan membolehkan model yang dibangun untuk pengecaman set muka yang tidak diketahui atau yang bara. Keistimewaan sistem ini adalah, ia dapat mengenali muka yang tidak diketahui dalam masa nyata. Sistem ini menggunakan algoritma Viola-Paul untuk mengesan muka, PCA dan LDA untuk mengekstrak data yang utama daripada muka, dan technique *Approximate Nearest Neighbour* untuk mengenal pasti muka tidak sama ada kenal atau tidak kenal melalui jarak metrik. Apabila muka tidak diketahui dikenal pasti, ia akan dipelajari dan dilabelkan oleh sistem. Oleh itu apabila muka yang sama muncul kali kedua, ia akan diiktiraf. Didapati, 60% ketepatan diperolehi selepas pengubahsuaian hyperparameter pemilih ciri dan jarak ambang metrik dilakukan.

Keywords: pengesanan muka masa sebenar, pengekstrakasi ciri muka, anggaran jiran terdekat, pengenalan muka secara set terbuka

CHAPTER 1

INTRODUCTION

Background of the Study

Nowadays, real-time face detection and recognition have gone viral and committed in the various daily application, like crowd identification, security footage analysis, security verification etc. This topic has brought to researcher attention due to the reason that human face is a dynamic object factor like pose, movement, and angle of face brought up to the camera could affect the capability of detection and human face also subject to a very high degree of variability in their appearance, which make face recognition a difficult problem in computer vision. Thus, most of the face recognition or face verification programme are required to run in the strictly controlled environment.

However, using controlled environment can only making a closed world assumption. Real-time recognition always touches upon open set classification with incomplete knowledge and too many unknown inputs (Scheirer, Jain, & Boult, 2014). As noted by research, when a when a recognition system is trained, there are a finite set of known objects in scenes with myriad unknown objects, combinations, and configurations. Labeling something new, novel or unknown should always be a valid outcome (Scheirer, De Rezende Rocha, Sapkota, & Boult, 2013).

Current machine learning formalized the problem solving to be closed set, where all the category are known in prior. Thus, to study this problem the data which not in training example should be used in testing, this is so called zero-shot learning (Lampert, Nickisch, & Harmeling, 2014). The algorithms presented in this project are Viola-Jones algorithm (Haar Cascade Classifier), PCA (classified as either feature based and image based), LDA (classified as either

feature based and image based), Approximate Nearest Neighbour and are discussed in terms of technical approach and how it is used to solve Open Set Recognition Problem.

A good approach to vision recognition must be able to deal with large database and could matching within seconds before fugitive has a chance to run away.

Problem Statement

Despite the world nature is dynamic and open, most of the recognition systems presume that the world is static and using a closed world model which treat every object must be known class in prior. However, in face recognition, we do not have the entire set of all possible faces. Thus, using closed world model of solving recognition problem is no longer useful in the way, the recognition system should be able to continuously update additional face category and accurate in detecting unknown or unseen face category.

The underlying problem of this recognition problem is due to the existing algorithms and method used for recognition possess weak generalization in handling variation in the human face, and unknown face frequently is treated as known category which will lead the recognition to wrong result (Scheirer et al., 2013).

This project formularizes a generalized method for recognition toward open set recognition approach. The proposed approach enable the model to recognize an infinite set of known faces in myriad of unknown faces and will treat unknown, unseen, new or novel faces to be a valid outcome, hence via the ability to determine unknown face, it will lead to a correct result, since the world is open and human faces are too many to be finite set to use in training a recognition system.

This leads to what is sometimes called "open set" recognition, in comparison to systems that make closed world assumptions or use "closed set" evaluation (Scheirer et al., 2013).

General Objective

The general objective for proposing a real-time open set recognition system is to design and develop a real-time open set facial recognition system that can recognize the previous known face and identify new, unseen or unknown face for learning it.

Specific Objectives

The specific objectives of this project are as stated:

(i) In-depth study of the various type of real-time recognition algorithm (Chapter 2).

(ii) To deploy real-time recognition with open source Python programming (Chapter 3).

(iii) To develop a functional real-time open set facial recognition system with integrating of facial detection, facial feature extraction, and machine learning classifier (Chapter 4).

(iv) To discuss the result and finding of real-time open set recognition system (Chapter 5).

Significance of the Study

The contribution of this study is helping researchers to make a precise and robust open set real-time recognition system. Besides that, this study can transfer the knowledge and experience of open set recognition to the next study. This technology can be helpful for the future individual and groups who work with open set recognition problems such as researcher, IT expert and industry.

In term contribution to the society, the security measurement in term of public safety will be improved if the technology proposed is applied in wide range of surveillance system, it capable of acting as artificial god eye which able to track anyone in sight of camera surveillance.

With the correct use of this system it could fight terrorism, find fugitives and etc, society will get benefit from this research where the crime rate might go down.

This proposed approach are potential to solve open set recognition problem. Consequently, this system will narrow down the literacy achievement gap with the current approach and the approach proposed by this project.

Scope of the Study

The real-time open set facial recognition system shall detect human faces and recognize face identity in real time. The system also expected to label the detected faces on the screen with their identity name. When an unknown face is captured the system shall save their image into image gallery, calibrate the classifier and label the face with new identity name eg: Person_3. Thus, the system can recognize the unknown when the unknown face appears again. This will enable the system to recognize an infinite set of the face.

In order to develop an open set facial recognition system, the proposed classifier in this system are using features rather than class to doing classification. This is due to the reason we have no all possible face for the class label, however, all human faces have unique facial features to describe their face. Thus, the approximate nearest neighbor is the best tool to classified the face identity according to the distance metric calculate from the human face feature extracted from PCA and LDA.

As the system are a lack of face tracking function, thus, it will not able to save and calibrate multiple new faces. Besides, the subject posture and illumination are the factors that will affect the accuracy of detecting unknown.

Definition of terms

Real-time. According to the definition provided by Oxford University Press (2017), the conceptual definition of real time is defined by the actual time when the process or event occurs. In computational term, real time is defined as the system that interprets the input signal and feedback immediately to the process from which is coming within a millisecond.

The operational definition of real time in this project refer to the process when the high frame rate information present in a video streaming is directly interpreted as an input signal and the system will immediately recognize the human object detected and feedback with the precise label of either known or unknown within a millisecond to screen of the device.

Open Set. According to Free Dictionary (2017), the conceptual definition of an open set is an interval on the real line excluding its end points, as [0, 1], the set of reals between, but excluding, 0 and 1.

The operational definition of an open set in this project refers to that the recognition is not presumed that there is exactly known the class of training, where the known class are in a myriad of unknown. Open set recognition can be further illustrated by, recognition is not just recognizing the population on the small island, but should take unknown or unseen data into consideration. When intrusion happen and your system wrongly recognized the unseen intruder as known person, how danger the situation it will become. Thus, open set recognition is open to the real world and not enclosed to a finite class of output assume that unknown will appear in testing rather than training phase.

Facial Recognition. According to the definition provided by Oxford University Press (2017), conceptual definition of facial recognition is stated as computer-aided identification of human faces by means of visible face characteristic and feature for security propose.

The operational definition of facial recognition is this study defined exactly similar to above conceptual definition. This project will utilize PCA (Eigenface) and LDA (Fisherface) to extract common facial characteristic and feature in the human face and use it for recognition and classification.

Chapter Summary

This chapter includes the background of the study, problem statement, general objective, specific objectives and significance of the study. The scope of study and definition of terms also discussed in this chapter.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter focuses on the literature review regarding the previous study of the component used in real-time facial recognition system and the open set recognition problem. The literature review divided into two aspects, such as related study and related work. The formal part studies the method of the real-time face detection algorithm, feature extraction model, approximate nearest neighbor classifier and the evaluate of the system. The latter part studies the similar system proposed from another researcher.

Related Study

Real Time Face Detection (Haar Cascade). The intuition of effective face object detection using HAAR feature based cascade classifier is first proposed by Paul Viola and Michael Jones (2001), this machine learning approach throw light into solving object detection problem. It utilizes Cascade function for training a large amount of binary labeled relevant images, which intent to be used for detection of the object. Figure 1 shows the 14 types of Haar feature which will recognize and identify from an image.

The advantage of HAAR feature based cascade classifier is that it could be used for detection for a wide range of object depending on the training data which it's fed. As an example to work out a face detector, we need to train the classifier with a lot of positive label face (images with a face) and negative label face (image without a face). Then it will generate HAAR feature as shown in Figure 1. A single frame will generate a pool of HAAR feature, then a very fast calculation method the integral image is used to allows for very fast feature evaluation, where it run convolution summation with kernel 2X2 (Viola & Jones, 2001).

However integral might not be the best approach since, in the pool of many features, there will exist not importance and irrelevant feature, the best way to work out has proposed on another paper. According to, they are using a classifier to discriminate every HAAR feature and check for the error of misclassifying the face. Only the feature with small error will be selected. The classifier will keep update with new feature and adjust its weight for minimizing error until the optimal error rate is found.

In the final stage, the Adaboost classifier is used to combine multiple weak classifiers to predict the relevant of the feature will further reduce the feature, and increase the accuracy of face detection. Rani and Muneeswaran (2014), presented a robust and effective face detection in real time video using Haar feature based with incorporate with Local Normalization technique to alleviate illumination variation.

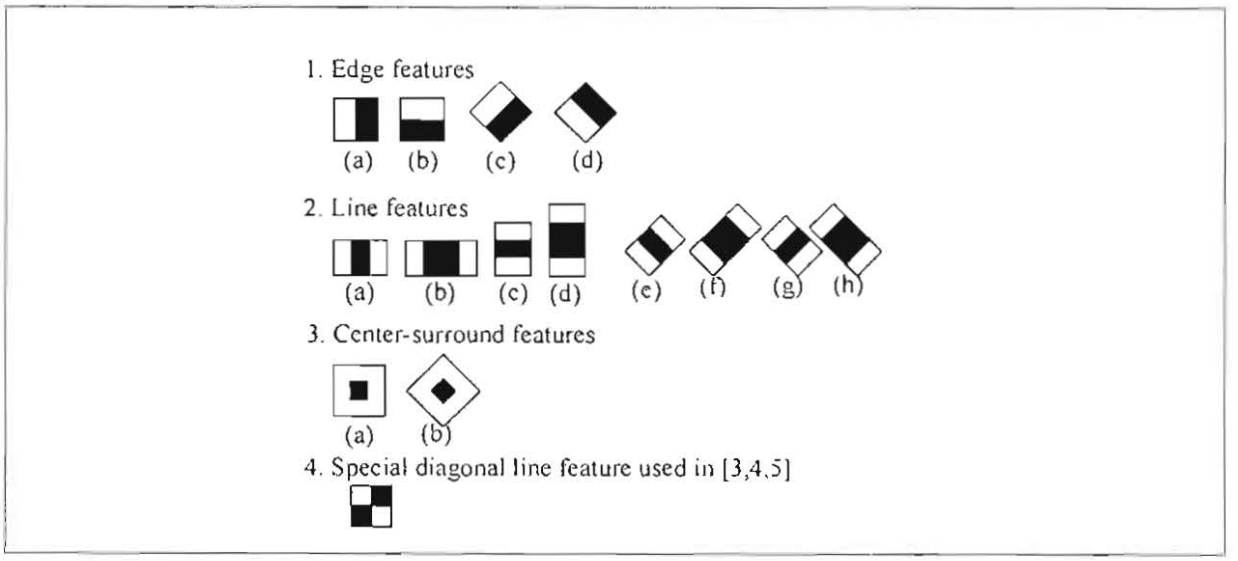

Figure 1. The 14 types of HAAR feature which will select from an image.

Feature Extraction. Dimensional Reduction and feature extraction is the component that always in used in supervised and non-supervised learning in classification and regression. Feature extraction is the subset of the feature is extracted and hence able to reduce the dimension of the matrix used in training. Besides, It extracts the meaningful feature subset from original dates by some rules, to reduce the time of machine training and the complexity of space (Hidayat, Fajrian, Muda, Huoy, & Ahmad, 2011). When dealing with data or classification problem, the large amount of feature will have scattered the space and make it complex and occupied. Besides that, some features are noise feature which with exacerbating the risk of overfitting. This is often when dealing with big data. The more the number of features, the higher the dimensionality.

According to a literature from Powell (2007), he describes that dimensionality is a curse to data science. Besides, dimensionality curse in supervised learning might be causing the training result to overfit (Zainal, Maarof and Shamsuddin, 2008). On the unsupervised learning side, such as clustering, the large amount of dimensionality will cause the definition of density or distance between point to be average out and become uniform (Janeck, Gangsterer, Demel& Ecker, 2008). Both cases show us that, high dimensionality will to low classification accuracy. The dimension of data must be reduced to decrease memory usage and decrease the cost of running those data. A study by Tan et al. (2005) support that reduces the data dimensional can help us to better visualize and understand the dataset.

Principle Component Analysis and Linear Discrimination Analysis are both linear transformation technique that transforms data feature to another space and remove excessive feature hence serves as a commonly used dimensionality reduction technic. PCA is an "unsupervised" algorithm because it does not look at the class labels to distinct each feature, whereas it just analyzes it principle component to its goal of finding maximum variance. LDA

are so-called "supervised" algorithms because whenever it computes the direction of axes that will maximize the separability of the classes, it needs to take input classes as a reference.

A research from Swets and Weng showed results of linear discrimination can capture discriminatory information better than using PCA as shown in figure 2. The figure 2(a) shown the sample of Principle component analysis have the tendency of the principal components to capture major variations in the training set such as lighting direction; Figure 2(b) shown the corresponding sample of Linear Discrimination Analysis has the ability of Linear Discrimination Analysis which can discount those factors unrelated to classification (Swets & Weng,1999).

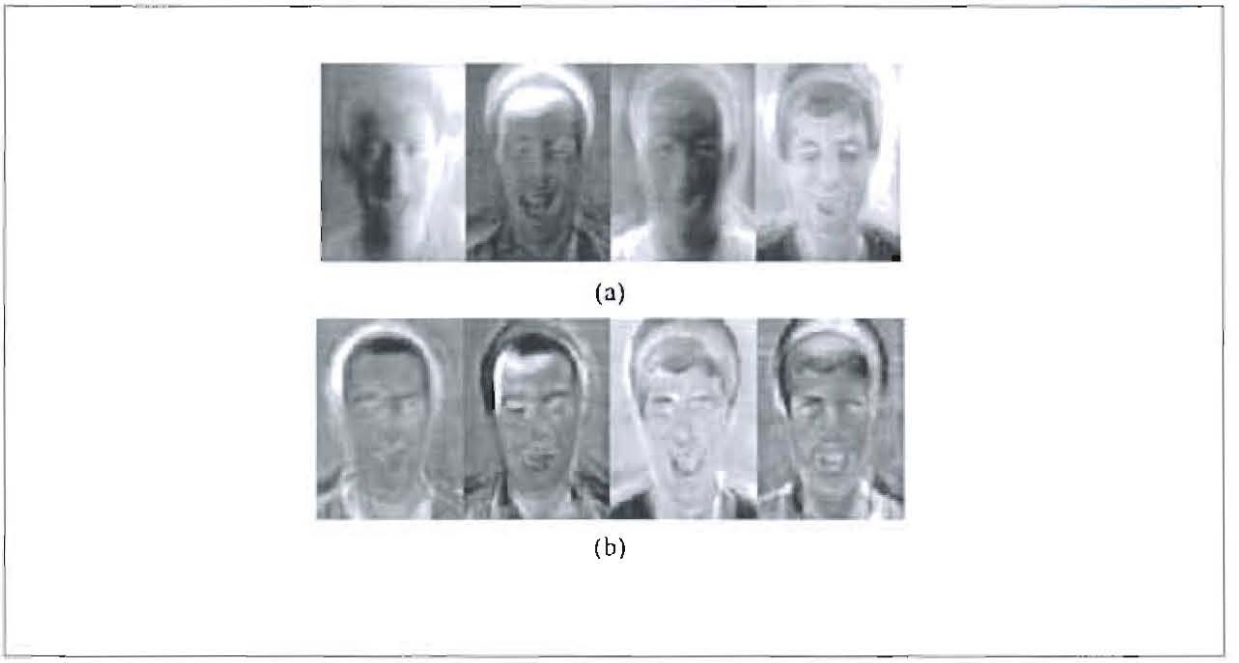

Figure 2. Comparison of feature capture by LDA and PCA.

PCA. Principle Component Analysis is commonly used dimensionality reduction technique for classification, data mining, pattern recognition, data compression and imaging problem (Song, Guo, & Mei, 2010). When training data is subject to very high dimensional space it will trigger dimensionality curse which causing classifier to overfit it.

Often, features are similar or correlated to the nearby pixel. Thus, PCA is a useful pre-processing technique to remove least discriminative or less statistically independent feature vector, hence maximize the variance (Li, Zhao, & Zhang, 2009). Besides, PCA only extracts the smallest number of the component that accounts for the variance for the data set and it able to summarize the whole dataset using the smallest number of principle component with little loss of information (Labib & Vemuri, 2006).

Before eliminating the feature, all the feature will be transformed into a feature space and those feature with similar covariance will be eliminated. While in image processing, the face image in PCA is considered as a random vector. PCA are intended to solving scatter matrix eigenvalue by samples a group of the new orthogonal base. These orthogonal bases are obtained to indicate the subspace spanned by training samples, and the features extracted are just project-vectors of the face images (Li et al., 2009). After performing PCA transformation, image X will be mapped to a lower dimension y as see Equation 1. In Equation 1, the image X is mapped to a lower dimension y.

$$y = W^T X \tag{I}$$

Where $X \in R^N, y \in R^M \ (N > M), W$ indicate projection matrix.

As stated on the above formula the number of example image M must be higher than the principle component N. Thus, a large number of the image should be involving into training for the PCA to select it invariance from the dataset provided. PCA require the number of

principal components is less than or equal to the number of original variables (Jain & Shandliya, 2013).

Covariance matrix S can be calculated using Equation 2:

$$S = \frac{1}{n}\sum_{i=1}^{m}(x_i - \bar{x})(x_i - \bar{x}) \quad (2)$$

Eigenvectors e_i can be formed using Equation 3:

$$(S - \lambda_i I)e_i = 0 \quad (3)$$

PCA-based feature extraction exploits a number of eigenvectors, by using eigenvectors corresponding to the first largest m eigenvalues to evaluate the significance of different feature components of the original samples (Song et al., 2010). By following rationale from the covariance matrix, the eigenvector corresponding to a larger eigenvalue can capture more representative information of an images (Xu, Zhang, & Yang, 2010).

PCA is good in maximize variance. On the other hand, PCA is also useful in the case to handle illumination variation compared to LDA approach (Cho, & Moon, 2009). However, due to PCA tendency in ignoring class separability, it might not perform well when the classes increase (Martinez & Kak, 2001). Without considering the class separability is the primary downside of PCA (Hidayat et al., 2011).

LDA. Linear discriminate analysis sometimes also known as Fisher's Discriminant Analysis (FDA). It is a common used dimensional reduction technique to simplify data by searching of optimal reconstruction of the sample and maximizing the separability between classes (Hidayat et al., 2011).

The key difference between principal component analysis and LDA is that the former deals with unlabelled data and tries to maximize variance, whereas the latter deals with labeled data and tries to maximize the discrimination between the classes (Zaki & Meira, n.d.).The formula to calculate the within-class scatter matrix sees Equation 4 and between-class see Equation 5 scatter matrix. In Equation 4, S_w denoted as within-class scatter matrix, $X_k^{(i)}$ is image of class $k(k = 1,2,3 \dots M)$. $\bar{x}_k$denoted the mean vector of class k. In Equation 5, S_B denoted as within-class scatter matrix, $\bar{x}_k$denoted the mean vector of class k. $\bar{x}$ is overall mean vector. $N_k$is the sample size of class k.

$$S_w = \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{N_k}[(X_k^{(i)} - \bar{x}_k)\left(X_k^{(i)} - \bar{x}_k\right)^T] \tag{4}$$

$$S_B = \frac{1}{M}\sum_{k=1}^{M}(N_k(\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})^T) \tag{5}$$

Given labeled data consisting of d-dimensional points x_i along with their classes y_i, the goal of linear discriminant analysis (LDA) is to find a vector w that maximizes the separation between the classes after projection onto w (Zaki & Meira, n.d.). Equation for getting W LDA optimal projection matrix form is defined in Equation 6.

$$J(W) = \frac{\|W^T S_B W\|}{\|W^T S_W W\|} \tag{6}$$